**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pylab **as** plt

**from** matplotlib **import** pyplot

df **=** pd**.**read\_csv('autodata.csv')

print("---------------Information---------------\n")

print(df**.**info())

print("\n")

print("---------------Describe the Dataframe---------------")

print(df**.**describe())

print("\n")

print("---------------First 10 rows---------------")

print(df**.**head(10))

print("\n")

print("---------------Last 5 rows---------------")

print(df**.**tail())

print("\n")

print("=================Data Preprocessing======================\n")

print("---------------Count of Null Values from the Dataset---------------")

print(df**.**isnull()**.**sum())

print("\n")

print("---------------Calculate the mean value for 'stroke' column---------------")

mean\_stroke **=** df["stroke"]**.**astype("float")**.**mean(axis**=**0)

print("Average with NULL values : ", mean\_stroke)

print("\n")

print("---------------Replacing NULL values with mean for 'stroke' column---------------")

df["stroke"] **=** df["stroke"]**.**replace(np**.**nan, mean\_stroke)

print(df["stroke"])

print("\n")

print("---------------Calculate the mean value for 'horsepower' column---------------")

mean\_horsepower **=** df["horsepower"]**.**astype("float")**.**mean(axis**=**0)

print("Average : ", mean\_horsepower)

print("\n")

print("---------------Replacing NULL values with mean for 'horsepower' column---------------")

df["horsepower"] **=** df["horsepower"]**.**replace(np**.**nan, mean\_horsepower)

print(df["horsepower"])

print("\n")

print("---------------Filling values in num-of-doors column with mode---------------")

mode **=** df["num-of-doors"]**.**mode()[0]

print("Mode of the Column : ", mode)

df["num-of-doors"] **=** df["num-of-doors"]**.**replace(np**.**nan, mode)

print(df["num-of-doors"])

print("\n")

print("---------------Dropping Null value rows---------------")

df**.**dropna(subset**=**['horsepower-binned'], axis**=**0, inplace**=True**)

df**.**reset\_index(drop**=True**)

print("number of null rows after : ", df['horsepower-binned']**.**isnull()**.**sum())

print("\n")

print("---------------Null value count after Preprocessing---------------")

print(df**.**isnull()**.**sum())

print("\n")

print("=================Data Standardization======================\n")

print("---------------Standardizing 'city-mpg' column to 'city-L/100km'--------------------")

df["city-L/100km"] **=** 235**/**df["city-mpg"]

print(df["city-L/100km"]**.**head())

print("\n")

print("---------------Standardizing 'highway-mpg' column to 'highway-L/100km'--------------------")

df["highway-L/100km"] **=** 235**/**df["highway-mpg"]

print(df["highway-L/100km"]**.**head())

print("\n")

print("=================Data Normalization======================\n")

print("---------------Normalizing 'length', 'width', 'height' column--------------------")

df['length']**=**df['length']**/**df['length']**.**max()

df['width']**=**df['width']**/**df['width']**.**max()

df['height']**=**df['height']**/**df['height']**.**max()

print(df[['length','width','height']]**.**head())

print("\n")

print("--------------turning categorical values into quantitative (numeric) variables--------------------")

print(df['aspiration']**.**value\_counts())

dummy\_var\_1**=**pd**.**get\_dummies(df['aspiration'])

print(dummy\_var\_1**.**head())

df**=**pd**.**concat([df,dummy\_var\_1], axis**=**1)

df**.**drop('aspiration',axis **=** 1 , inplace **=** **True**)

print(df**.**head())

print("=================Data Binning======================\n")

print("---------------Binning 'horsepower' column into bins--------------------")

df["horsepower"]**=**df["horsepower"]**.**astype(float, copy**=True**)

plt**.**hist(df["horsepower"])

plt**.**xlabel("horsepower")

plt**.**ylabel("count")

plt**.**title("horsepower bins")

plt**.**show()

print("\n")

print("Horsepower binned")

bins **=** np**.**linspace(min(df["horsepower"]), max(df["horsepower"]), 4)

group\_names **=** ['Low', 'Medium', 'High']

df['horsepower-binned'] **=** pd**.**cut(df['horsepower'], bins, labels**=**group\_names, include\_lowest**=True** )

print(df[['horsepower','horsepower-binned']]**.**head(20))

pyplot**.**bar(group\_names, df["horsepower-binned"]**.**value\_counts())

plt**.**xlabel("horsepower")

plt**.**ylabel("count")

plt**.**title("horsepower bins")

plt**.**show()